

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-117020

(43)Date of publication of application : 19.04.2002

(51)Int.Cl.

G06F 17/21  
G06F 12/00

(21)Application number : 2000-311114

(71)Applicant : NEC CORP

(22)Date of filing : 11.10.2000

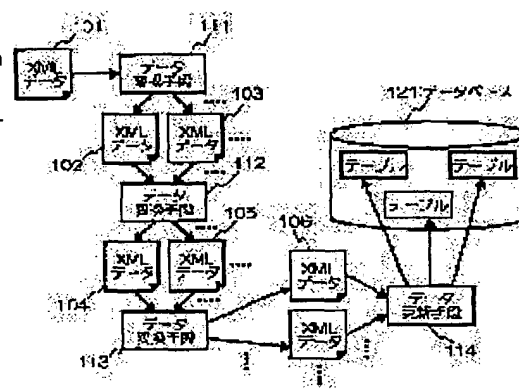
(72)Inventor : NAKAMURA HIDEO

## (54) DEVICE AND METHOD FOR STORING DATA

## (57)Abstract:

PROBLEM TO BE SOLVED: To store an XML document, including a nesting structure, a repetitive, or a recursive structure of elements (tags) in  $\geq 3$  layers, in a table type database.

SOLUTION: This data storage device which converts XML data 101 into table type data and stores them in the database 121 is equipped with a data converting means 111 which finds and separates an internal repetition structure included in the XML data under a new root element, a data converting means 112 which converts a recursive tag structure in the XML data into a repetition structure, a data converting means 113 which converts low-order elements included in a multi-layered nesting structure in the XML data into high-order elements, and a data updating means 114 which stores the converted into elements divisionally in tables corresponding to respective new root elements.



## LEGAL STATUS

[Date of request for examination]

19.09.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

特開 2002-117020  
(P 2002-117020 A)  
(43) 公開日 平成 14 年 4 月 19 日 (2002. 4. 19)

(51) Int. Cl. <sup>7</sup>	G 0 6 F	17/21	識別記号	5 7 0	FI	17/21	5 7 0 G 58009	5 7 0 L 58082	5 0 1 T	5 4 8 Z	12/00	5 1 3 A	密査請求 有	請求項の数 7	OL	(全 13 頁) 最終頁に続く

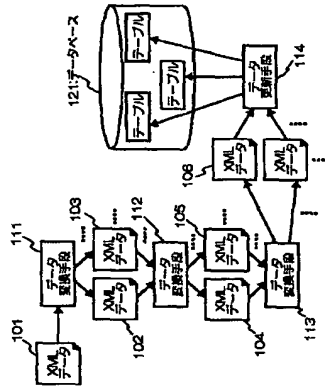
(21) 出願番号	特願 2000-311114 (P2000-311114)	(71) 出願人	000004237 日本電気株式会社
(22) 出願日	平成 12 年 10 月 11 日 (2000. 10. 11)	(72) 発明者	中村 秀男 東京都港区芝五丁目 7 番 1 号
		(74) 代理人	100108578 井理士 高橋 昭男 (外 3 名) F ターム (参考) 58009 NAO5 S407 58082 G402

(54) 【発明の名称】 データ格納装置及び方法

(57) 【要約】

【課題】 3 階層以上の要素 (タグ) の入れ子構造、繰り返し、または再帰的構造を含む XML 文書を、テーブル形式のデータベースに格納することができるようにする。

【解決手段】 XML データ 101 を表形式のデータに変換してデータベース 121 に格納するデータ格納装置であって、XML データに含まれる内部の繰り返し構造を括り出して新たなルート要素の下に分離するデータ変換手段 111 と、XML データに含まれる再帰的なタグ構造を繰り返し構造に変換するデータ変換手段 112 と、XML データに複数階層の入れ子構造で含まれる下位の要素を上位の要素へ繰り上げる変換を行うデータ変換手段 113 と、変換された要素を各ルート要素に対応する複数のテーブルに分割して格納するデータ更新手段 114 とを備えている。



【特許請求の範囲】

【請求項 1】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

変換すべき XML データに複数階層の入れ子構造で含まれる下位の要素を上位の要素へ繰り上げる変換を行う変換手段と、  
変換された要素をテーブルに格納する格納手段とを備えることを特徴とするデータ格納装置。

【請求項 2】 前記変換すべき XML データが 3 階層以上の入れ子構造を有していることを特徴とする請求項 1 のデータ格納装置。

【請求項 3】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

変換すべき XML データに含まれる内部の繰り返し構造を括り出して新たなルート要素の下に分離する変換手段と、

変換された要素を各ルート要素に対応する複数のテーブルに分割して格納する格納手段とを備えることを特徴とするデータ格納装置。

【請求項 4】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

変換すべき XML データに含まれる再帰的なタグ構造を繰り返し構造に変換する変換手段と、  
変換された要素をテーブルに格納する格納手段とを備えることを特徴とするデータ格納装置。

【請求項 5】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

変換すべき XML データに含まれる要素の位置を示す情報を作成するとともに、その XML データに入れ子構造で含まれる下位の要素を上位の要素へ繰り上げる変換を行う変換手段と、

変換された要素を、位置対応情報とともにテーブルに格納する格納手段とを備えることを特徴とするデータ格納装置。

【請求項 6】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

XML データに複数階層の入れ子構造で含まれる下位の要素を上位の要素へ繰り上げる変換を行う第 1 の変換手段と、

XML データに含まれる内部の繰り返し構造を括り出して新たなルート要素の下に分離する第 2 の変換手段と、  
XML データに含まれる再帰的なタグ構造を繰り返し構造に変換する第 3 の変換手段と、

変換された要素を各ルート要素に対応する複数のテーブルに分割して格納する格納手段とを備えることを特徴と

するデータ格納装置。

【請求項 7】 XML (eXtensible Markup Language) データを表形式のデータに変換して所定の記憶手段に格納するデータ格納装置において、

変換すべき XML データに複数階層の入れ子構造で含まれる下位の要素を上位の要素へ繰り上げる変換を行う第 1 の変換手段と、

変換すべき XML データに含まれる内部の繰り返し構造を括り出して新たなルート要素の下に分離する第 2 の変換手段と、

変換すべき XML データに含まれる再帰的なタグ構造を繰り返し構造に変換する第 3 の変換手段と、

変換された要素を各ルート要素に対応する複数のテーブルに分割して格納する格納手段とを有していることを特徴とするデータ格納装置。

【発明の詳細な説明】

(0001)

【発明の属する技術分野】 本発明は、要素 (タグ) の入れ子構造、繰り返し、再帰的構造を含む XML (eXtensible Markup Language) 文書をテーブル形式のデータベースに格納する際に用いて好適なデータ格納装置及び方法に関する。

(0002)

【従来の技術】 XML (eXtensible Markup Language : 拡張可能なマークアップ言語) は、W 3 C (World Wide Web Consortium) で標準化が進められている Web 上で構造化文書をやりとりするためのデータフォーマットである。XML は、SGML (Standard Generalized Markup Language) [ISO 8879] の利用を前提とし、そのサブセットとして設計されている。XML 文書は、ルートと呼ばれる文書全体から始まり、それぞれマークアップ付けられ、かつ入れ子構造を有している宣言、要素、コメント、文字参照および処理命令を含んでいる。各文書は一つ以上の要素を含み、各文書にはルートまたは文書要素という要素が一つだけ存在し、これは他の要素の内容に含まれない。各要素は、開始タグと終了タグで区切られ、入れ子構造をなしている。すべての要素は、その開始タグが他の要素の内容に含まれれば、対応する終了タグも同じ要素の内容に含まれる。また、要素 B が要素 A の内容に含まれ、かつ要素 A の内容に含まれる他の要素が含まれないとき、要素 A を要素 B の親といい、要素 B を要素 A の子という。

(0003) 図 18 を参照して、XML 文書を構成するデータ (以下、XML データ) をテーブル形式のデータベースに格納するための XML データ格納方式について説明する。図 18 において、データ格納手段 401 は、XML データ 421 の内容を解析し、XML データ 421 内に含まれる各要素を抽出してデータベース 402 内のテーブル 422 に格納する。XML データ 421 が図 19 に示すような内容を有している場合を例にして、図

(0004) 図 18 を参照して、XML 文書を構成するデータ (以下、XML データ) をテーブル形式のデータベースに格納するための XML データ格納方式について説明する。図 18 において、データ格納手段 401 は、XML データ 421 の内容を解析し、XML データ 421 内に含まれる各要素を抽出してデータベース 402 内のテーブル 422 に格納する。XML データ 421 が図 19 に示すような内容を有している場合を例にして、図

(0005) 図 18 を参照して、XML 文書を構成するデータ (以下、XML データ) をテーブル形式のデータベースに格納するための XML データ格納方式について説明する。図 18 において、データ格納手段 401 は、XML データ 421 の内容を解析し、XML データ 421 内に含まれる各要素を抽出してデータベース 402 内のテーブル 422 に格納する。XML データ 421 が図 19 に示すような内容を有している場合を例にして、図

18を示すデータ格納手段401の動作について説明する。この場合、XM1データ421は1つのルート要素421aを持ち、要素421bの子要素として、次の1つ下の階層に要素421b(421b1、421b2)の樹形返しを持っている。各要素421b1、421b2は、要素421c、d、eからなる同じ子要素の列をそれぞれ含んでいる。データ格納手段401はXM1データ421を入力として受け取り、要素412aの行各子要素412b1、412b2をテーブル422の行B01、B02に対応させて格納する。また各子要素412b1、412b2の子要素412c-eを各行の列C、D、Eに対応させて格納する。テーブル422の例を図20に示す。

1000

【発明が解決しようとする課題】上述したような従来技術には、次のような問題点があった。第1の問題点は、要素421 aの子要素421 b、2の子要素421 c、要素421 aの子要素421 b1、2の子要素421 b2、eがさらに子要素を持つ場合にテーブルの列に対応させて格納することができないという点である。その理由は、テーブルの列が要素421 aの子要素421 b1、2の子要素421 c～eに対応しているためである。第2の問題点は、要素421 aの子要素421 b1、2の内部に繰り返しを持つ場合にテーブルに格納することができないという点である。その理由は、テーブルの行が表現する繰り返しが要素421 aの子要素にだけ対応しているためである。第3の問題点は、再帰的なタグ構造をテーブルに格納できないということである。その理由は、テーブルが再帰的なデータ構造をそのまま格納できないためである。

【0005】本発明は、従来の構成では対応できなかった、3階層以上の要素（タグ）の入れ子構造、繰り返し、または再帰的構造を含むXML文書を、テーブル形式のデータベースに格納することができるようにするデータ格納装置及び方法を提供することを目的とする。

[0006]

【問題を解決するための手段】 上記問題を解決するた

め、請求項1記載の発明は、XMLデータを変換形式のデータに変換して所定の処理手段に格納するデータ格納装置において、変換すべきXMLデータに複数階層の入れ子構造で含まれた下位の要素と上位の要素と階層の入れ子構造を行なう変換手段と、変換された要素をデータに格納する格納手段とを備えることを特徴とする。請求項2記載の発明は、前記変換すべきXMLデータが3階層以上の入れ子構造を有していることを特徴とする。

【０００７】請求項３記載の発明は、XMLデータを表形式的データに変換して所定の記憶手段に格納するデータ格納装置において、変換すべきXMLデータに含まれる内部の繰り返し構造を齎り出して新たなルート要素の下下下に分離する変換手段と、変換された要素を各ルート要素に対応する複数のデータプールに分割して格納する格納手

データ105およびその他のXMLデータと、分離した部分を除いたXMLデータ104となる。XMLデータ103等、データ変換手段1110のその他の出力XMLデータもXMLデータ102と同様の変換を行う。データ変換手段1113はXMLデータ104の文書定義上3レベル以上の要素の入子になっている要素をルート要素の子要素の子要素とする形式に変換しXMLデータ106とする。XMLデータ105等、データ変換手段1122のその他の出力XMLデータ105についても同様に変換を行う。データ変換手段1111~113の変換によりXMLデータ101はルート要素の子要素としてそれぞれ同じ要素の繰り返しを待ち子要素の内部に同じ要素の列を待たされる。データ更新手段1114はこれらのXMLデータ106およびその他のXMLデータ105にこれらのXMLデータ106およびその他のXMLデータ105をデータベース121のルート要素の子要素をテーブルの列に対応付けることによってXMLデータをデータベース121のテーブルに格納する。このようにして、3レベル以上のXMLタグの入れ子構造、繰り返し、再帰的タグ構造を含むXML文書をデータベース形式のデータベースに格納することとを可能にする。

【0011】次:

格料装置の他の変換形態について説明する。図2を参照すると、本装置のXMLデータ格料装置の一実施形態では、XMLデータ入力手段201と、データ変換手段202と、データ変換手段203と、データ変換手段204と、データ格料手段205と、データベース211から構成されている。XMLデータ221の例を図3に示す。XMLデータ221はルート要素A（開始タグ：<A>、終了タグ：</A>）以下構成。ただし図面では開始タグに参照符号を引き出し線を付けている。1から構成されている。要素Aは要素B（B01、B02）の繰り返しで構成されている。要素B01は要素C（C01）と、要素D（D01、D02）から構成されている。要素B02は要素C（C02）と、要素D（D03）から構成されている。要素C01は1つの要素E（E01）と、要素F（F01、F02）の繰り返しから構成されている。要素C02は1つの要素E（E02）と、要素F（F03）から構成されている。要素D01は1つの要素G（G01）と1個の要素D（D02）から構成されている。要素D03は1つの要素G（G03）から構成されている。要素F（F01、F02、F03）はそれぞれ要素Iと要素Jから構成されている。要素E、要素G、要素I、要素Jは文字列から構成されている。

【0012】上記各手段はそれぞれ概略つぎのように動作する。データ入力手段201は、XMLデータ221を入力し、データ変換手段202へ渡す。データ変換手段202は、XMLデータ221の内部の繰り返し要素である要素F(F01,F02)を分離しXMLデータ222とXMLデータ223とする。要素B(B01,B

02) も繰り返しているが、ルート要素Aの子要素なので変換しない。データ変換手段203は、XMLデータ222の内部の再帰になっている要素D(D01、D02)を分離しXMLデータ224とXMLデータ225とす。データ変換手段204は、XMLデータ224の内部の3レベル以上入予になっている要素Eをルート要素Bの子要素として移動しXMLデータ226とす。データ格納手段205は、XMLデータ226をテーブルAに格納する。このとき要素Lの内容を列Eに格納する。XMLデータ225をテーブルYに格納する。このとき要素Gの内容を列Gに格納する。XMLデータ223をテーブルXに格納する。このとき要素Iの内容を列Iに、要素Jの内容を列Jに格納する。

【0013】[実施形態の動作の説明]次に、図4、図5、図6、図7のフローチャートを参照して本実施形態の全体の動作について詳細に説明する。

【0014】XMLデータ221をデータ変換手段202へ入力する。要素Xをルート要素とする新しいXML

20 データ223を生成する(図4のステップA101)。

ここでXは整数である。XMLデータ22.2.1に要素Bがなければ終わり(ステップA102)、要素Bがあれば、最初の要素Bを要素B1とする(ステップA103)。

ここで、BおよびB1は整数を示す。次に、要素B1に要素Fがなければ終わり(ステップA104)、要素B1に要素Fがあれば要素B1の最初の要素FをF1とする(ステップA105)。ここでFおよびF1は整数である。要素F1を要素B1から取り除き要素Xに追加する(ステップA106)、要素B1に次の要素Xがあれば

30 はステップA108へ、なければステップA109へ進む(ステップA107)。ステップA108では次の要素FをF1として、ステップA106へ戻る。ステップA109では、次の要素BがあればステップA110へ、なければ終わる。ステップA110では、次の要素BをB1として、ステップA104へ戻る。終了時には総ての要素Fが取り除かれたXMLデータ221をXMLデータ222とする。

【0015】データ交換手段203はデータ交換手段202からXMLデータ2222を入力する。変数Yをルート要素とする新しいXMLデータ2225を生成する(図5のステップA201)。ここでYは変数である。XMLデータ2222に要素Bがなければ終わり(ステップA202)、要素Bがあれば最初の要素Bを要素B1とする(ステップA203)。次に要素B1の子要素の要素DをD1とする(ステップA204)。ここでDおよびD1は変数である。次に要素D1を親要素から取り除き、変数Yの子要素として追加する(ステップA205)。要素D1の子要素に要素DがあればステップA207へ、なければステップA208へ進む(ステップA206)。



11  
1とする(ステップA507)。要素XにY1を追加する(ステップA508)。CT1をB1から取り除き要素Y1に追加する(ステップA509)。要素Y1に要素IDを追加する(ステップA510)。要素IDの内訳としてID1を生成する(ステップA511)。要素B1に次の要素Cか文字列があればステップA513へ、なければステップA514へ進む(ステップA515)。次の要素Cか文字列をCT1とする(ステップA516)。次の要素BがあればステップA515へ、なければステップA514へ進む(ステップA515)。次の要素BをB1に追加する(ステップA514)。次の要素Bの子の総ての要素Fと文字列を取り除かれたXMLデータ321をXMLデータ322とする。

10  
[0026] データ格納手段303はデータ変換手段302からXMLデータ323を入力する。XMLデータ323に要素Yが含まれていない(図15のステップA601)。XMLデータ323の最初の要素Yを要素Y1とする(ステップA602)。テーブルXに行X1を追加する(ステップA603)。要素Y1に要素Cがなければ内容を行X1の列Cに格納する(ステップA604)。要素Y1に文字列があれば行X1の列A、A605。要素Y1に文字列があれば行X1の列TEXTに格納する(ステップA606、A607)。要素Y1の子要素IDの内容を列X1の要素IDに格納する(ステップA608)。次の要素BがあればステップA610へ、なければ終わる。次の要素BをB1とする(ステップA610)。次に、具体例について説明する。

19  
[0027] XMLデータ321をデータ変換手段302へ入力する。要素Xをルート要素とする新しいXMLデータ323を生成する(図14のステップA501)。最初の要素B、要素B01を取り出す(ステップA503)。要素B01の順序位置IをID1とする(ステップA504)。要素B01の最初の文字列CT01を取り出す(ステップA506)。要素Yを生成しY01とする(ステップA507)。要素XにY01を追加する(ステップA508)。CT01をB01から取り除き要素Y01に追加する(ステップA509)。要素Y01に要素IDを追加する(ステップA510)。

20  
[0028] データ格納手段303はデータ変換手段302からXMLデータ323を入力する。XMLデータ323の最初の要素Y、要素Y01を取り出す(図15のステップA602)。テーブルXに行X01を追加する(ステップA603)。要素Y01の文字列を行X01の列TEXTに格納する(ステップA607)。要素Y01の子要素IDの内容を行X01の列IDに格納する(ステップA608)。次の要素Y02を取り出す(ステップA610)。テーブルXにY02を追加する(ステップA603)。要素Y02の子要素IDの内容を行X02の列IDに格納する(ステップA608)。次の要素Y03を取り出す(ステップA610)。テーブルXにY03を追加する(ステップA603)。要素Y03の文字列を行X03の列TEXTに格納する(ステップA607)。要素Y03の子要素IDの内容を行X03の列IDに格納する(ステップA608)。次の要素Y04を取り出す(ステップA610)。テーブルXにY04を追加する(ステップA603)。要素Y04の子要素Cの内容を行X04の列Cに格納する

13  
(ステップA605)。要素Y04の子要素IDの内容を行X04の列IDに格納する(ステップA608)。次の要素Y、要素Y05を取り出す(ステップA610)。テーブルXに行X05を追加する(ステップA603)。要素Y05の文字列を行X05の列TEXTに格納する(ステップA607)。要素Y05の子要素IDの内容を列X05の列IDに格納する(ステップA608)。次の要素Y、要素Y06を取り出す(ステップA610)。テーブルXに行X06を追加する(ステップA603)。要素Y06の子要素Cの内容を行X06の列Cに格納する(ステップA605)。要素Y06の子要素IDの内容を行X06の列IDに格納する(ステップA608)。作成されたテーブルXを図17に示す。

14  
[0029] 本実施形態の第1の効果は、文字列、要素が混在した繰り返し行がテーブル形式のデータベースに格納できることである。その理由は、タグと文字列の繰り返し構造を持つ部分と要素あるいは文字列と要素として保持し格納できる形に変換したためである。第2の効果は、テーブル形式のデータベースに格納された文字列、要素のデータの元の位置情報を格納できることにある。その理由は、文字列や要素を格納する際に、その情報もテーブルに格納したためである。

15  
[0030] なお、本発明の実施の形態は上述した形態に限定されるものではなく適宜変更可能である。例えば、各実施形態における変換手段の位置は他の変換手段と交換可能である。図2を参照して説明した実施形態に、図18を参照して説明した実施形態で用いた要素の位置情報を要素を追加する構成を付加することなどが考えられる。また、本発明のデータ格納装置は、コンピュータとそのコンピュータで実行されるプログラムとを用いて実現することができ、そのコンピュータで実行されるプログラムはコンピュータ読み取り可能な記録媒体あるいは通信回線を介して頒布することが可能である。

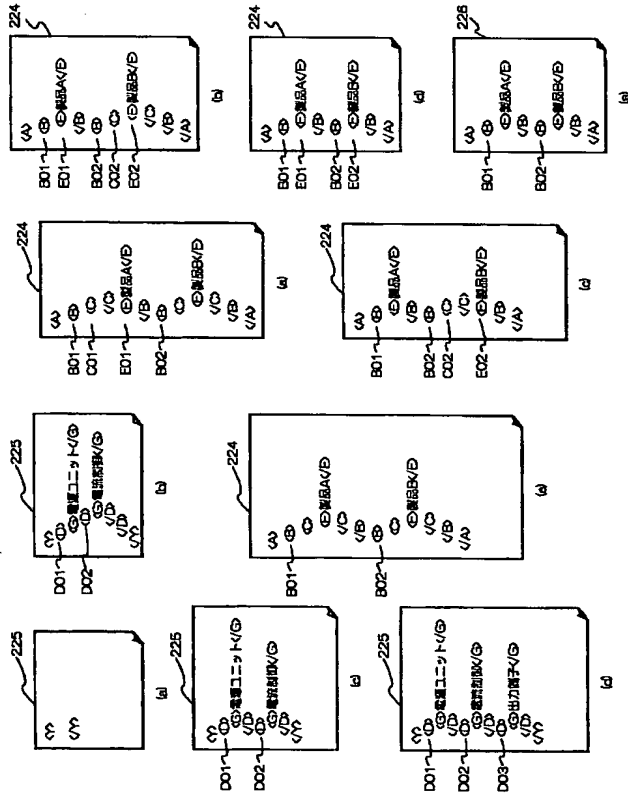
16  
[発明の効果] 以上説明したように発明によれば、次のような効果を得ることができる。第1の効果は、タグの繰り返し構造を含むXML文書をテーブル形式のデータベースに格納できることにある。その理由は、タグの繰り返し構造を持つ部分を別のデータとして格納し、テーブル形式に格納できる形に変換したためである。第2の効果は、再帰的タグ構造を含むXML文書をテーブル形式のデータベースに格納できることにある。その理由は、再帰的タグ構造を持つ部分を別のデータとして格納し、テーブル形式に格納できる形に変換したためである。第3の効果は、3レベル以上のタグ入りデータ構造を含むXML文書をテーブル形式のデータベースに格納でき

17  
ることにある。その理由は、3レベル以上の入れ子構造を持つ部分の要素を上位の要素へ移動しテーブル形式のデータベースに格納できる形に変換したためである。第4の効果は、文字列、要素が混在した繰り返し行がテーブル形式のデータベースに格納できることである。その理由は、タグと文字列の繰り返し構造を持つ部分と要素あるいは文字列と要素として保持し格納できる形に変換したためである。第5の効果は、テーブル形式のデータベースに格納された文字列、要素のデータの元の位置情報を格納できることにある。その理由は、文字列や要素を格納する際に、その情報もテーブルに格納したためである。

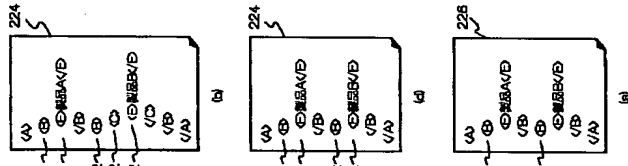
18  
[図面の簡単な説明]  
[図1] 本発明によるデータ格納装置の一実施の形態の構成を示すブロック図。  
[図2] 本発明によるデータ格納装置の一実施の形態の構成を示すブロック図。  
[図3] 図2のXMLデータ221の一例を示す図。  
[図4] 図2の構成(データ変換手段202)の動作を示すフローチャート。  
[図5] 図2の構成(データ変換手段203)の動作を示すフローチャート。  
[図6] 図2の構成(データ変換手段204)の動作を示すフローチャート。  
[図7] 図2の構成(データ格納手段205)の動作を示すフローチャート。  
[図8] 図2におけるデータ変換手段202によるXMLデータ223の作成経過(a)～(d)と作成されたXMLデータ222(e)を示す図。  
[図9] 図2におけるデータ変換手段203によるXMLデータ225の作成経過(a)～(d)と作成されたXMLデータ224(e)を示す図。  
[図10] 図2におけるデータ変換手段204によるXMLデータ224の作成経過(a)～(d)と作成されたXMLデータ226(e)を示す図。  
[図11] 図2におけるデータ格納手段205によるデータベース211内のテーブルA、X、Yの作成経過(a)～(m)を示す図。  
[図12] 本発明によるデータ格納装置の他の実施の形態の構成を示すブロック図。  
[図13] 図12のXMLデータ321の一例を示す図。  
[図14] 図12の構成(データ変換手段302)の動作を示すフローチャート。  
[図15] 図12の構成(データ格納手段303)の動作を示すフローチャート。  
[図16] 図12においてデータ変換手段302によって作成されたXMLデータ323の内容の一例を示す図。



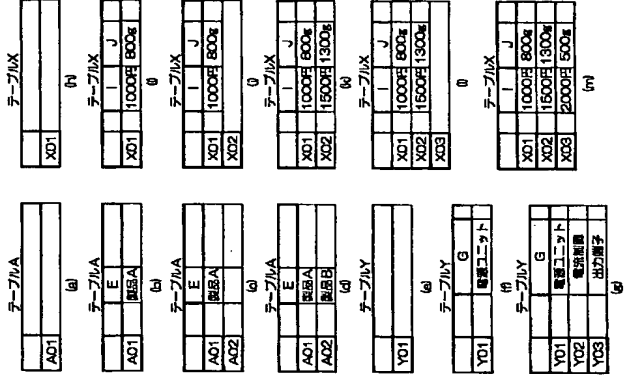
【图9】



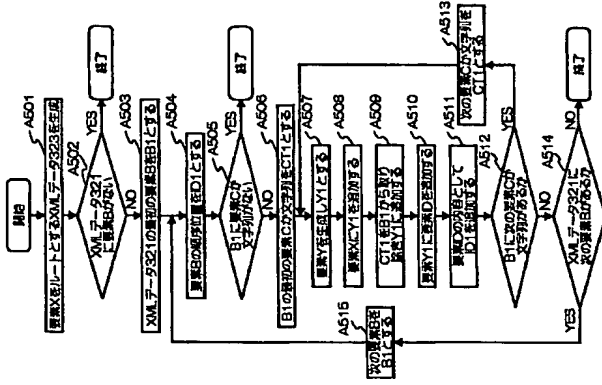
【圖 10】



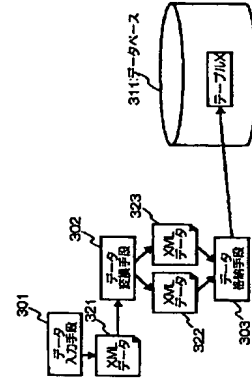
【图 11】



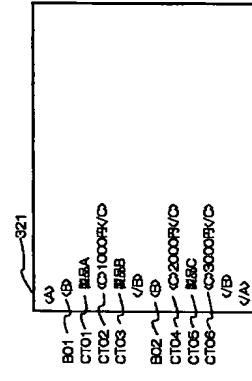
【图14】



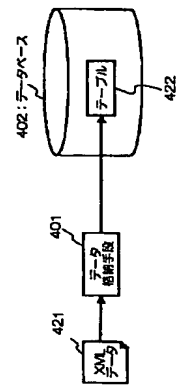
【图 12】



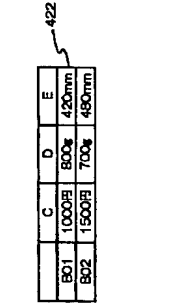
【圖 13】



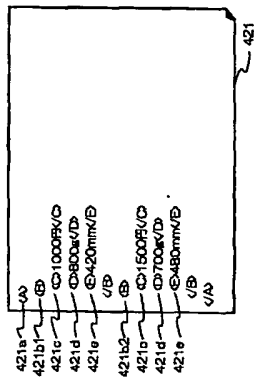
【图 18】



【图20】



【図19】



フロントページの続き

(51) Int. Cl.<sup>7</sup> G 0 6 F 1 2 / 0 0 識別記号 5 2 0 F I G 0 6 F 1 2 / 0 0 5 2 0 J (参考)